Selecting Public Goods Institutions: Who Likes to Punish and Reward?

Drouvelis, Michalis; Jamison, Julian C.

DOI:
10.1002/soej.12049

Document Version
Peer reviewed version

Citation for published version (Harvard):
Drouvelis, M & Jamison, JC 2015, 'Selecting Public Goods Institutions: Who Likes to Punish and Reward?', *Southern Economic Journal*. https://doi.org/10.1002/soej.12049

Selecting Public Goods Institutions: Who Likes to Punish and Reward?*

Michalis Drouvelis
University of Birmingham
Email: m.drouvelis@bham.ac.uk

Julian C. Jamison
Consumer Financial Protection Bureau
Email: julison@gmail.com

July 2014

Abstract: We study the link between individual attitudes toward uncertainty on the one hand, and preferences over, as well as behavior within, various public goods institutions on the other hand. In particular, we measure individual levels of risk aversion and ambiguity aversion over both gains and losses. We then incentive-compatibly elicit preferences over voluntary contribution mechanisms with and without reward and punishment options. Finally we randomly assign subjects to one of the four institutions and observe repeated play. We find that payoffs are significantly greater when punishment is allowed but that only a small minority of participants prefers such an environment. Somewhat surprisingly, preferences over institutions are generally independent of individual characteristics. On the other hand, institutional preferences, as well as individual characteristics, are significantly predictive of behavior in the public goods game. For instance, risk averse individuals preemptively punish more often when that option is available. This result suggests that when studying social behaviors involving sanctions and rewards, it is important to consider individual attitudes toward risk and uncertainty—although they may not affect the original selection into institutions.

Keywords: public goods; voluntary contribution; risk, loss, and ambiguity aversion; preference elicitation; reward and punishment
JEL codes: C92; H41; D81

* We thank Jeff Butler, Antonio Cabrales, Urs Fischbacher, Simon Gächter, Brit Grosskopf, Zack Grossman, John Hey, the editor and two anonymous referees, and audiences at the Southern Economic Association meetings, the 2012 IMEBE conference, and the Universities of Cyprus and East Anglia for helpful comments; Lynn Conell-Price for research assistance; Walter Yuan for programming assistance; and the University of York for financial assistance. Remaining errors are due solely to the authors. The views expressed are those of the authors and do not necessarily reflect those of the Consumer Financial Protection Bureau or the United States.

I. Introduction

Institutions are an integral part of our social life and organize important aspects of economic activity. It is well-established that different institutions lead to different economic outcomes (North 1990). How institutions are defined also has serious implications for the evolution of human culture and societies (Tabellini, 2008). For example, Acemoglu, Johnson, and Robinson (2001, 2002) show that different colonization strategies and policies have led to differences in the institutions implemented in the respective colonized countries, which consequently have affected their long-run economic welfare. In related work, by using a large sample of countries Easterly and Levine (2003) and La Porta et al. (1998) demonstrate that institutions can explain differences in the levels of a country's economic development and financial performance, respectively. Thus, our understanding of the link between institutional structures and outcomes, as well as what individual-level factors are predictive of this process, is of great interest for economists and social scientists.

In this paper, we use experimental methods to shed more light on these questions and to gain a better understanding of how the underlying institutions enforce a society's norms. We focus on institutions that are concerned with providing public goods in settings where free-riding incentives are present. The general motivation for studying these institutions stems from the fact that a number of real-life situations (for example, tax compliance, donations to charities, and participation in collective actions) are characterized by an incentive structure where people's individual and collective goals are at odds. In addition, experimental behavior in these institutions has inspired the recent development of novel theoretical models of social preferences (see Camerer 2003; Fehr and Schmidt, 2006) which account for a number of the observed anomalies. Therefore, identifying which forces determine the content of acceptable standards of behavior captured by these institutions will shed further light on the proximate sources of human cooperation.

A more specific motivation for our study comes from a burgeoning experimental literature that investigates individuals' voting preferences over public good institutions and over the specific rules that govern these institutions. A main message from these studies (see the related literature section for a more extensive review) suggests mixed evidence on which institutions people favor, but in general it is observed that democratically selected institutions (by using a certain voting rule) perform better than institutions that have been exogenously imposed, both in terms of average contribution levels and efficiencies measured by net earnings. Yet the existing literature does not address two important issues in settings

involving public goods. First, given the likelihood of strategic concerns when voting, which institutions do individuals actually prefer? Second, what individual characteristics have predictive power over their preferred institutional choice and of behaviors in the relevant institutions?

Our experiment utilizes an incentive-compatible mechanism (which also has applicability in other domains) to elicit preferences over a menu of four institutions, each with different enforcement mechanisms to punish and/or reward behavior that does not comply with a certain norm: a standard voluntary contributions mechanism (VCM), a VCM with opportunities to sanction, a VCM with opportunities to reward, and a VCM with opportunities to sanction and reward. By observing how individuals select institutions from the available set of options, we are able to draw conclusions about which enforcement mechanism is actually preferred; see details below. By observing how much individuals are prepared to pay or how much they would need to be paid to participate in a given institution, we elicit the intensity of their preferences.

To the best of our knowledge, our paper is the first experimental study which elicits preferences regarding public good institutions and the intensity of these preferences in an incentive-compatible manner.[1] Our study also contributes to the understanding of social norms by simultaneously analyzing the three enforcement mechanisms that have played a central role in the social preferences literature. As a control case, we include an institution where neither sanctions nor rewards are present. The comparison of behavior among the three institutions using different enforcement mechanisms enables us to disentangle which particular institutional aspect (sanctions and/or rewards) of an institution, if any, is important for sustaining norms of high cooperation and maximizing individuals' overall welfare. We believe that the use of laboratory experiments is ideal for addressing these questions, as collecting data on our variables of interest is often infeasible in naturally occurring environments. Additionally, incentivizing preferences and assessing efficiency issues is typically difficult in the field due to a number of factors that operate simultaneously, confounding the analysis of causal relationships.

[1] Technically we need to be a bit careful about the exact usage of incentive-compatibility in this paper. The ordinal rankings are indeed incentive-compatible, and the cardinal rankings (intensities) have the right relative ratios within an individual, but may be compressed in absolute levels due to risk aversion. Cardinal preferences are always hard to compare across individuals, and that is even more true here due to potential differences in risk attitudes. In practice, as reported below, we do not see any empirical relationship between risk aversion and the absolute value of expressed intensities (not surprisingly given the stakes), so this absence is unlikely to cause any issues when interpreting the results.

Furthermore, the level of tight control provided by experimental methods allows us to elicit a number of variables that we hypothesize may influence subjects' choice of institutions and subsequent play. Our central focus is on preference measures related to uncertainty, such as risk, loss, and ambiguity aversion. Partly this is because previous work has tended to focus on the social preference dimension in the context of public goods: altruism, reciprocity, etc. While that is undoubtedly important, it leaves out central factors having to do with risk and payoff range, which are especially relevant in more complex (and realistic) settings when punishment and reward are allowed. For example, institutions with punishment options have the potential to be detrimental (or at least to add variability) to subjects' welfare and thus may be preferred by risk-seeking subjects. In contrast, institutions with reward mechanisms may be preferred by individuals who are averse to ambiguity. Also, loss averse and risk averse individuals may be attracted to institutions where sanctions and rewards are not available.

Our decision to focus on attitudes toward uncertainty is driven not just by the conceptual arguments above, but also by previous studies that found such a link in other institutional settings. Laboratory studies such as Bartling et al. (2009) and Dohmen and Falk (2011) have found that subjects who are less risk-averse are more likely to sort into competitive environments or variable-pay schemes. Similarly, Weinhold and Zak (2005) find that risk attitudes are central to wage-related occupational choice in China. Finally and perhaps most relevantly, there is a relatively long history in the contract theory literature studying the link between risk tolerance and contract choice, going back at least to Cheung (1969). A striking recent example is Ackerberg and Botticini (2002), who examine agricultural institutions in Renaissance Italy and show that risk-sharing is a major determinant of contract choice.

Although the papers described above are suggestive, the relationship between risk tolerance and social preferences, including voluntary contributions, has received relatively scant direct attention. There have been a few experimental studies (for example, Eckel and Wilson 2004; Humphrey and Renner 2011; Kocher et al. 2011), that have yielded equivocal evidence. Importantly, all these studies overlook the possibility that preferences other than standard risk preferences, such as loss and ambiguity aversion, may predict social preferences. These studies also do not credibly elicit preferences over specific social mechanisms for procuring public goods. In our paper, we elicit choices in order to construct four preference measures (risk aversion, loss aversion, ambiguity aversion, and ambiguity aversion over losses) in an incentive-compatible manner, and we provide the first

comprehensive analysis of how these preference measures can predict subjects' reciprocal behavior in the form of punishing or rewarding their peers.

Our main findings can be summarized as follows. First, our four preference measures are significantly correlated with each other. Second, which institutions individuals prefer are, surprisingly, not influenced by preference measures, although other individual traits do have some explanatory power. Third, institutions with punishment options are best able to maintain cooperative norms. Fourth, relative to institutions without sanctioning mechanisms, institutions that permit sanctions incur enforcement costs that lower overall welfare in the short run but increase overall efficiency in the long run. Fifth, positive and negative reciprocity are significantly correlated with our preference measures. Sixth, subjects' individual characteristics account for the way sanctions and rewards are used.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 outlines our experimental design and section 4 presents the results from our data analysis. Section 5 concludes by discussing the implications of our results and how these might be extended by further work.

II. Related Literature

There has been extensive exploration of how people behave in decision situations where a tension exists between personal and collective gains. By now it is well-understood that individuals contribute part of their endowment but contributions gradually decline as the game progresses (see Ledyard 1995; Gächter and Herrmann 2005). Many economists have sought to identify mechanisms in order to remedy the free-rider problem, and focus on public good environments where individuals have the opportunity to punish and/or reward their peer group members (see Chaudhuri 2011 and Gächter and Herrmann 2009).[2] Our paper contributes to the existing literature which examines behavioral dynamics in public good games when individuals are given the opportunity to exhibit preferences over their environment.

This literature has mainly focused on the selection of environments that offer rewards and punishments. In a study by Botelho et al. (2005) subjects were asked to vote for their preferred environment after they acquired experience by playing for 10 periods a standard

[2] In disciplines other than economics, the implications of punishment have also received considerable attention (see, e.g., Yamagishi 1986; Ostrom, Walker, and Gardner 1992; de Quervain et al. 2004).

public good game and a public good game with sanctioning opportunities. After the voting took place, the majority of votes determined which environment all subjects played in for a final period. This study found that subjects did not favor the sanctioning environment and also found that sanctions did not have a sustained positive effect on contributions and profits.

Margreiter, Sutter and Dittrich (2005) examined common pool resource (CPR) games, where subjects could propose appropriation levels for each member of their group and vote on the proposals put forward. In some of their sessions, they had heterogeneous groups and in other sessions, they had homogeneous groups.[3] Their findings indicate that homogeneous groups implement a proposal more often and have a higher overall efficiency than heterogeneous groups. Another experiment by Gürerk, Irlenbusch, and Rockenbach (2006) had subjects choose at the beginning of every period whether they would like to play in an environment without or with sanctions. In each period, a participant then interacted with all the other participants who had chosen the same institution. Their results provide evidence that given a "voting-with-one's-feet" approach the sanctioning environment becomes the predominant choice over time.[4]

Sutter, Haigner, and Kocher (2010) examined how endogenous selection affected three public goods environments (a standard public good game, a public good game with punishment, and a public good game with rewards) by letting subjects vote whether or not to accept each of the available environments. When the cost-to-impact ratio of rewarding was 1:3, 85 percent of the groups agreed on the public good game with reward, whereas with a ratio of 1:1, only 25 percent did. In their endogenous treatments, Sutter, Haigner, and Kocher find that the reward environments (both under a 1:1 and a 1:3 ratio) and the punishment environment (under a 1:1 ratio) generate higher contribution levels than the standard public good game.[5] A number of recent experimental studies have further indicated that democratically selected institutions have positive effects on behavior.[6] Ertan, Page, and Putterman (2009) find that institutional environments where subjects vote on whom they are allowed to punish yield higher contributions and greater efficiencies than institutions in which punishment is unrestricted. Noussair and Tan (2011) extend this study by allowing the existence of heterogeneity in the value that subjects' contributions generated for the group

[3] In the heterogeneous groups, half of the subjects in a group had a high cost of appropriating one token from the CPR and the other half had a low cost, and in homogeneous groups, all group members had the same cost.
[4] In a later paper, Gürerk, Irlenbusch, and Rockenbach (2009) provide further evidence for this finding.
[5] The finding that punishment has a larger effect on contribution behavior and efficiency when democratic institutions are available is also in line with Tyran and Feld (2006).
[6] For a review of the experimental literature on the workings of democratic institutions, see Dal Bó (2010).

and find that the most effective institution (in terms of contributions and earnings) is the one that allows punishment of below-average contributors only. In another study, Putterman, Tyran, and Kamei (2010) provide evidence that giving subjects the opportunity to vote on the penalty structure in an environment that allows punishment will lead to efficiency-enhancing outcomes relative to an environment where subjects are not given this opportunity. Kamei (2011) also finds that when a sanctioning policy is implemented democratically, subjects who favor the policy contribute more to their group relative to when the policy is implemented exogenously. In another series of experiments, Dal Bó, Foster, and Putterman (2010) also show that letting subjects democratically choose which policy they prefer positively affects cooperative behavior.

The existing literature suggests that voting is the prevalent procedure used to implement individuals' preferences over environments. Although voting is a valid method for endogenously assigning individuals to specific environments, its disadvantage is that voting is not necessarily an incentive-compatible process. Our experimental investigation contributes to the "endogenous selection" literature in at least two respects. First, as described briefly in the introduction and in detail below, we introduce a fully incentive-compatible mechanism to elicit preferences over different public good environments. Second, our design consists of all four possible environments (a standard public good game, a public good game with punishment, a public good game with rewards, and a public good game with punishment and rewards), thus allowing us to compare their relative appeal as well as to disentangle the different individual and collective incentives that might be at work in these environments. Further, since we randomize assignments, we can separate the effect of selection from the effect of the institutional rules per se.

Given our interest in studying the link between individual attitudes toward uncertainty and preferences over public good environments, our paper contributes to existing literature that explores the interaction of preference measures and social preferences. A few experimental studies have recently addressed how preference measures interact with social preferences, but evidence from the available studies is mixed and focuses mainly on trust games. In a trust game in rural Paraguay, Schechter (2007) shows that risk aversion plays an important role in determining behavior, whereas Eckel and Wilson (2004) find no significant relationship between risk measures and the decision to trust. In another experiment, Kocher et al. (2011) provide evidence that there is no correlation between risk preferences and behavior in a public good game and in a trust game.

These experiments mainly address the issue of whether and how behavior in games measuring social preferences is affected by preference measures. While this is a relevant research question to identify the determinants of prosocial behavior, another important question largely ignored in this literature is what motivates the demand for different public good environments. This is a significant omission, as better understanding the self-selection process can help us improve the design of public institutions that promote social welfare and cooperation. By designing a novel experiment which elicits preferences over a menu of public good institutions in an incentive-compatible manner, our paper systematically investigates whether and how individual preferences affect the choice of institutions and subsequent behavior.

III. Experimental Design

Our experiment was conducted in two parts. In the first part we elicited subjects' levels of risk and ambiguity aversion (over both gains and losses). In the second part we elicited preferences over four public good environments and then randomly assigned subjects to one of these environments to play a repeated voluntary contribution game. At the beginning of the experiment, subjects were informed that the experiment would consist of two parts (in order to reduce the likelihood of incorrect expectations about the nature of the experiment). However, they were not told what would happen in the second part of the experiment (see Appendix A for the timeline of tasks that occurred in each session).

In order to elicit their preferences, participants were shown a table with seven rows and asked to choose in each row between a safe option and a lottery option. The safe option was exactly the same in each row, but the amount in the lottery option increased from one row to the next. More precisely, in the first row subjects could choose to receive £6 with certainty, or they could choose to play the lottery and have a 50 percent chance of receiving £0 and a 50 percent chance of receiving £11. Moving down the table, the amount it was possible to win in the lottery increased to £12, £13, £14, £16, £18, and £20. After a subject had made a decision for each row, it was randomly determined which row became relevant for payoff. Subjects were informed of their lottery payment at the end of the experiment. This procedure guaranteed that each decision was incentive-compatible. The number of times a subject chose

the safe option indicates his or her attitudes towards risk; that is, the more times a subject selected the sure payoff of £6, the more risk averse this subject is.[7]

As our public good environments involve payoffs that are ambiguous and may even involve losses, we consider it important to elicit individual attitudes towards loss aversion, ambiguity aversion, and ambiguity aversion with losses.[8] To elicit such preferences we implemented a procedure similar to the one used to elicit risk preferences. For instance, to elicit individual attitudes towards losses, we used the exact same table described above but with payoffs shifted downwards by £3. Thus the lottery payoffs now involved losses, as these consisted of a 50 percent chance of losing £3 and a 50 percent chance of receiving a positive amount. As a measure of loss aversion, we used the frequency with which a subject chose the safe option. For the cases of ambiguity aversion with and without losses, we simply replaced the probability of each outcome, made explicit in the lottery option, with a question mark to indicate that the probability was unknown. The four tables were shown to subjects in a random order to control for order effects. In particular, if the risk and loss questions (with 50–50 lotteries) always preceded the ambiguity questions, we might expect subjects to have a 50–50 prior distribution over outcomes when considering the ambiguous lotteries. The randomized presentation of these questions minimized this effect.

After the elicitation of preference measures, subjects received new instructions describing each public good environment (see Appendix B). In total, we examined four different environments, each corresponding to a separate treatment, with the individual participants experiencing only one treatment (a between-subjects design). We refer to our four treatments as: a) voluntary contributions mechanism (VCM); b) VCM with punishment; c) VCM with reward; and d) VCM with punishment and reward. In each session, a group of 12 subjects were randomly assigned across the above treatments to play a 25-period repeated game in groups of four. The group composition remained the same throughout the session (that is, we implemented a partner matching protocol). Earnings were given in money units for the public good games and we used an exchange rate of £0.01 per money unit.

[7] The simple number of sure-thing choices is the most natural and standard measure of risk-aversion, although they are not guaranteed to be in the first rows of the table. Of course, empirically we find (for loss and ambiguity as well) that subjects are more likely to take the gamble the higher is the relevant payoff.

[8] For control purposes we wanted to use a symmetric procedure to elicit individuals' preferences toward risk, loss and ambiguity. Thus we returned to the original Binswanger (1980) method of changing magnitudes rather than the Holt and Laury (2002) method of changing probabilities.

We will begin by describing our baseline treatment, VCM, and comment on the structure of the remaining treatments in turn.[9]

a) VCM treatment

Our baseline treatment is a stylized model that captures the conflict between private and social interests and is called the voluntary contributions mechanism (VCM) with linear payoffs. Under this treatment, subjects are randomly assigned to a four-person group and endowed with 20 tokens each. We used tokens rather than actual monetary units (as in the tasks above) in order to conform with the standard approach used in this literature. Each subject has to decide how much of this endowment to keep for themselves and how much to contribute to a public good (described to subjects as the "project"). For each token kept, a subject earned one money unit for themselves, while for each token contributed to the project, each of the four subjects in the group earned a return of 0.4 money units, resulting in a total of 1.6 money units for the whole group. After all group members made their contribution decisions, they were informed of the total amount of all contributions made to the public good and of their own income.

This simple baseline treatment allows us to measure the extent of self-interested behavior: since a subject's contribution cost one money unit, while the private return is only 0.4 money units, a selfish group member always has an economic incentive to contribute nothing to the public good and rely on the contribution of other group members. Yet social efficiency requires that all group members contribute their entire endowment to the public good (in this case each group member receives an income equal to 32 money units, which is greater than his/her initial endowment).

b) VCM with punishment treatment

The VCM with punishment treatment is identical to the VCM treatment except for the addition of a second stage. After subjects made their contribution decisions during the first stage, the other three group members' contribution profiles are revealed at the beginning of the second stage. No individual subject could identify the particular contribution of any other group member, since the order of contributions shown in each screenshot randomly changed from period to period, and therefore, subject-specific reputations could not develop across periods. Each subject could then assign between zero and five negative points to each of the

[9] In the actual instructions we used neutral framing in the description of the public good games. In particular, we referred to VCM as "Institution A," VCM with punishment as "Institution B," VCM with reward as "Institution C," and VCM with punishment and reward as "Institution D."

other group members.[10] Assigning negative points was costly both to the punisher and the punished group member, as each negative point costs the punisher one money unit and the punished group member three money units. At the end of the second stage, each subject was informed about the cost incurred for assigning negative points, the total number of negative points assigned to them, and their earnings from each period.

c) VCM with reward treatment

The VCM with reward treatment has a similar two-stage structure to the VCM with punishment treatment, except that each subject is given the opportunity to assign up to five positive points to other group members—assigning positive points is costly to the donor but beneficial for the recipient. Each positive point costs the donor one money unit and awards the recipient one money unit.[11] As in the previous treatments, subjects received information about their own rewards and earnings but group information was not provided.

d) VCM with punishment and reward treatment

The VCM with punishment and reward treatment is a combination of the VCM with punishment and the VCM with reward. After subjects made their contribution decisions in the first stage and their group's contribution profile was revealed, each subject was given the opportunity to assign either up to five negative or up to five positive points to each of the other group members. Thus, the overall earnings π_i of a group member i for a given period are as follows:

$$\begin{aligned} \pi_i &= 20 - g_i + 0.4 \cdot \sum_{j=1}^{n} g_j \\ &- \sum_{j \neq i} p_{ij} - 3\delta_{pun} \cdot \sum_{j \neq i} p^{pun}{}_{ji} + \delta_{rew} \cdot \sum_{j \neq i} p^{rew}{}_{ji}, \end{aligned} \tag{1}$$

where g_i denotes group member *i*'s contribution to the public good, p_{ij} denotes the number of points (either negative or positive) that group member *i* assigns to group member *j*, p_{ji} denotes the points assigned by *j* to *i*, either positive or negative according to the superscript, and finally the δ is a 0 or 1 indicator for whether punishment and/or reward is allowed.

[10] In the instructions, we refer to punishment as "assignment of negative points," and to reward as "assignment of positive points."

[11] We implement a 1:1 cost-to-impact ratio for the reward treatment to avoid creating a game of reciprocal creation of value in the rewarding process. In addition, the ratios of 1:1 for reward and 1:3 for punishment are commonly used in the previous literature.

Note that conditional on each subject i being motivated to maximize payoffs given by equation (1), the unique subgame perfect equilibrium for the VCM with punishment and reward requires that subjects free ride completely in the first stage and refrain completely from assigning points in the second stage. Of course, as with all previous literature in this area, this is not what we observe to occur, implying that subjects care about more than simply their expected monetary payoff.

As soon as subjects read the instructions for each treatment, they received a number of computerized control questions to ensure that they understood the decision situation and the payoff calculations. All participants had to answer these questions correctly; otherwise, the experiment would not proceed. Next, subjects were asked to indicate on a percentage scale how much they expected to earn relative to the maximum potential earnings, considering only earnings from the 25 rounds of the public good game (see figure C.1 in Appendix C for a screenshot of this step). We are interested in this assessment as a means to find out whether—and if so how—overconfidence affects the selection into public good games. Subjects received a bonus based on the accuracy of their estimation.[12] Subjects were informed about their true rank in the distribution at the end of the experiment.

After the subjects answered the overconfidence question, their institutional preferences were elicited. During this phase, the subjects were asked to indicate in which public good institution they preferred to participate by quantifying how much each institution was worth to them (in pounds and pence) overall.[13] The incentive system was as follows: subjects were asked to indicate a monetary amount for their preferred institution. They were told that if they were assigned to one of the institutions that they indicated they were willing to pay for, then the monetary amount they indicated would be subtracted (once) from their final payment. If they were assigned to an institution for which they had indicated that they would need to be paid to participate, then the monetary amount they stated would be added (once) to their final payment. Note that the maximum amount they could state was any number from –£5 to £5 (inclusive) and that the sum of all four amounts was required to be 0. To control for order effects, each institution appeared onscreen in a random order across participants. Figure C.2

[12] In particular, if their estimation was within 10 percentage points in either direction of their actual earnings (calculated as a percentage of the maximum earnings), they received £1; if their estimation was within 15 percentage points in either direction of their actual earnings, they received £0.50, and if their estimation was within 25 percentage points in either direction, they received £0.20.

[13] Of course they had not yet experienced any of the settings, but merely had them described. This allows us to examine their unvarnished attitudes to each environment. Sutter, Haigner, and Kocher (2010) make the same decision for the same reason: an interest in underlying institutional preference rather than individual learning.

in Appendix C provides a screenshot of the interface we used for eliciting subjects' preferences for each institution.

Our incentive mechanism allowed subjects to truthfully express the ordinal ranking of their preferred institution as well as the strength of their preference (by stating a monetary value for the amount that their preferred structure was worth to them). As long as subjects have diminishing marginal utility for money (that is, they would prefer to be given money in a low-income state of the world even if an equivalent amount were taken away in a high-income state), this mechanism induces them to smooth their income by stating that they prefer exactly those environments in which they expect to earn more, exactly like insurance.[14] It is therefore incentive compatible, and there are no relevant strategic considerations because the actual choice of institutions is completely random, which subjects are told in advance. That being said, probably the main reason it elicited useful information is simply because subjects have no reason not to tell us the truth: we are asking their preferences and they are responding, regardless of the monetary incentives. There is a tradeoff between using the ordinal rankings (which have better intra-subject comparability) and the cardinal values (which incorporate more information about intensity of preferences).[15]

After subjects entered their relative preferences, they were informed which of the four environments they had been assigned to and then played the game for 25 rounds. Although it removed any element of endogenous selection, we felt that randomization was important at least for this first iteration of the experiment. One rationale was purely logistical, ensuring that we had the right-sized groups within each session and that we got similar amounts of data for each possible treatment. Another rationale was to avoid any strategic manipulation in the elicitation mechanism, as described above. But most importantly, we wanted to see whether subjects who e.g. feared a particular institution (such as one allowing punishment) were correct to do so, which required that some of them actually face such an institution. Note that in the 'real world', both exogenous and endogenous institutional frameworks are quite common.

[14] On a related note, this elicitation procedure does add additional *ex ante* uncertainty to their payoffs, so one might worry that the magnitude of the expressed preferences was a function both of actual underlying intensity and of risk aversion. However, when we regressed the average absolute value of the expressed preferences against the normalized risk, loss, and ambiguity attitudes, we found no relationship—suggesting that this was not a concern in practice.

[15] We have analyzed the data with both approaches, and the main results are robust to this choice.

After the 25 rounds of play concluded, subjects were informed of their payoff from the lottery task, the overconfidence question (along with their actual rank in the distribution), and their earnings from the public good game. At this point we also collected data on the subjects' demographic characteristics (such as gender, age, nationality, marital status, father's education, political and religious affiliations) and on a self-control task that is correlated with cognitive outcomes (see Frederick 2005). Note that we did not elicit for a second time their preferences over institutions, although it would be interesting to see if and how it changed. Partly this was due to logistics (any incentive-compatible method for doing so would have required playing additional rounds of the game itself) and partly it was because theories of learning and belief-change were outside the scope of this investigation.

Procedures

We conducted sixteen sessions, four sessions for each of the four treatments. A total of 192 subjects participated in the experiment and in each of the four treatments there were 48 subjects. All the subjects were recruited at the University of York, using the ORSEE software (Greiner 2004). The vast majority of participants were undergraduate students from various academic fields. The experiment was conducted in the Centre for Experimental Economics (EXEC) lab and all treatments were computerized and programmed with the Multistage software from Caltech. The instructions for the elicitation of risk preferences and the description of the public good environments are provided in Appendix B. Some of the instructions were presented on the computer screen. At the end of a session, subjects were paid in private according to their total earnings from all relevant tasks. Average earnings per treatment were as follows: £13.54 for the VCM, £13.80 for the VCM with punishment, £14.19 for the VCM with reward, and £14.62 for the VCM with punishment and reward (at the time of the experiment £1 was equivalent to $1.61). Sessions lasted, on average, 70 minutes, with no session taking more than 90 minutes.

IV. Results

In the following three subsections, we present the main results from our experiment. In the first subsection, we focus on behavior in the public good games, both in contribution levels and in the assignment of points for punishment and rewards. After observing their actual behavior, we investigate how preference measures affect the subjects' choice of institutions.

Finally, we examine the responses that indicated the subjects' individual attitudes toward risk, loss, and ambiguity and how these preferences are interrelated and related to the subjects' responses in the CRT questions.

1. Behavior in institutions

1.1 Contribution behavior and net earnings

In this subsection, we discuss the results in the first stage of the VCM treatments. First we present our findings with respect to subjects' contribution behavior and then we analyze the efficiency of each of these institutions as measured by the subjects' average net earnings.

Figure 1 shows the average contributions made to the public good project across all 25 rounds, smoothed by a five-period moving average. We observe very similar patterns of average contributions between the standard VCM treatment and the VCM with reward treatment. In particular, subjects initially contribute approximately 50 percent of their total endowment, with group allocations declining to roughly 10 percent of the initial endowment after 25 rounds of play. However, as is apparent in Figure 1, the time trends and average contributions diverge when we examine the treatments which allow opportunities for punishment; that is, the VCM with punishment and the VCM with punishment and reward. Initial average contributions start from approximately the same point relative to the VCM and the VCM with reward treatments, but as the game progresses the contributions dramatically increase and move closely together. A visual inspection of Figure 1 suggests pronounced differences between the punishment and the no-punishment treatments, which are documented by our statistical analysis reported in Table 1.

Figure 1: Average contributions in each period by treatment

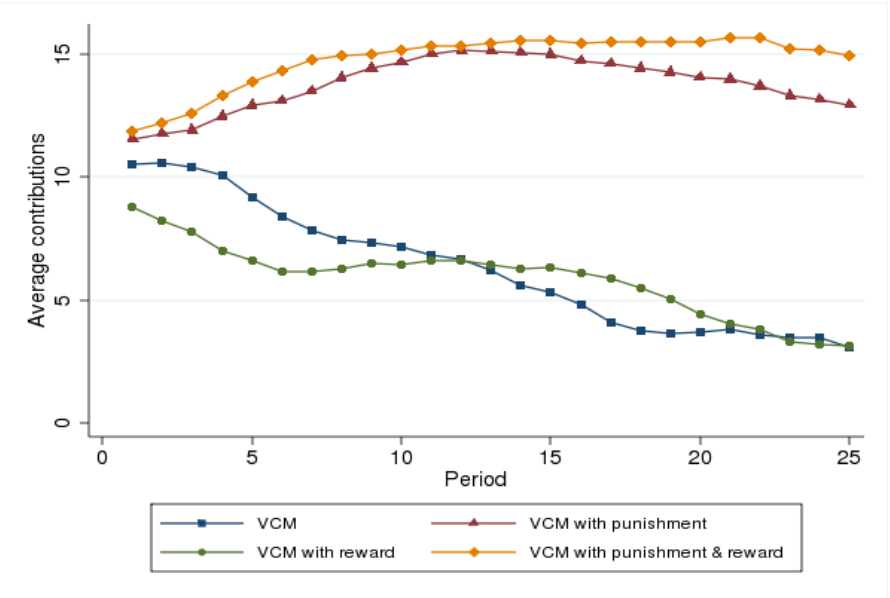


Table 1 presents the p-values of the non-parametric ranksum Wilcoxon test for each possible comparison between a pair of treatments. In parentheses, we also report the average contributions for each treatment across all 25 periods. In particular, we observe that the average contributions are largest in the VCM with punishment and reward treatment (14.73 tokens) and the VCM with punishment (13.74 tokens) and lowest in the VCM (6.25 tokens) and the VCM with reward treatments (5.86 tokens). Our statistical analysis records significant differences at the 1 percent level between the treatments with and without punishment.

Table 1. Average Contributions and p-Values of Pairwise Comparisons

	VCM (6.25)	VCM with punishment (13.74)	VCM with reward (5.86)	VCM with punishment & reward (14.73)
VCM (6.25)	--			
VCM with punishment (13.74)	0.00	--		
VCM with reward (5.86)	0.66	0.00	--	

VCM with punishment & reward (14.73)	0.00	0.73	0.00	--

Notes: Numbers in cells correspond to p-values for each pairwise comparison (using a ranksum Wilcoxon test). Numbers in parentheses correspond to average contributions in each treatment across all 25 periods.

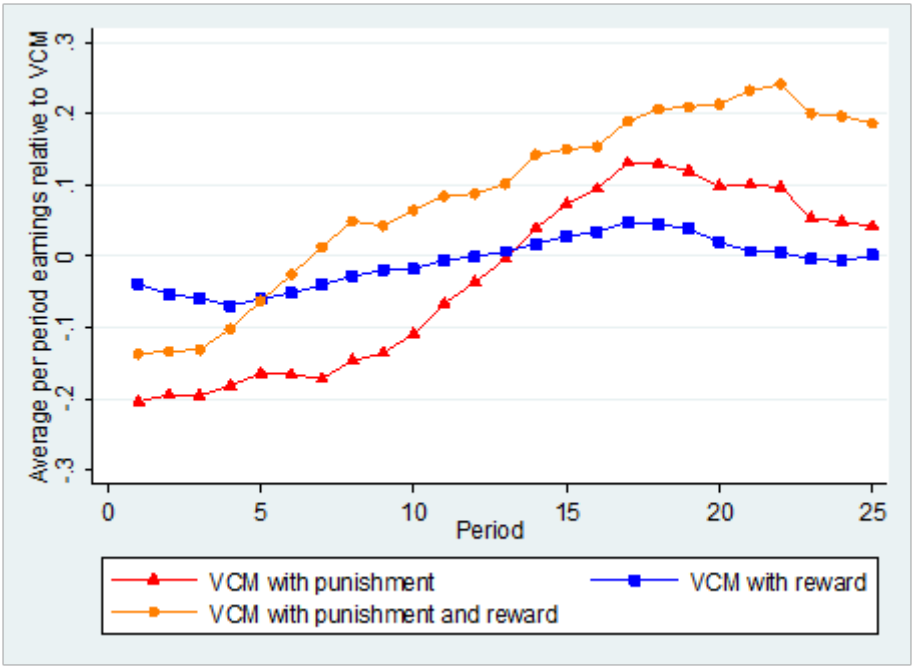
We further examine the efficacy of the treatments with and without punishment by looking at how average net earnings evolved over time. At the beginning of the game, we find that the VCMs with punishment yield lower average net earnings, but as the game progresses the average net earnings increase, resulting in higher average profits. The opposite pattern is observed in the treatments without punishment. The net profits are not significantly different between treatments with and without punishment when we average across all 25 periods (ranksum Wilcoxon test, $p = 0.503$; 24.19 tokens in the punishment treatments and 23.64 tokens in the no-punishment treatments). However, profits are lower ($p = 0.001$) in the first 5 periods of the punishment treatments (21.93 tokens) as compared to the ones with no punishment (25.45 tokens). Importantly, this trend is reversed across the final 10 periods, where average net profits are significantly higher ($p = 0.016$) for the punishment treatments (25.46 tokens) relative to the treatments with no punishment (22.41 tokens). Our findings provide further support for previous experimental evidence suggesting that the availability of a punishment mechanism decreases average net earnings in the short run, but causes an increase in efficiency in the long run (e.g., Gächter, Renner and Sefton, 2008).

To further test that it is the punishment opportunities that cause an increase in the efficiency of the public good institutions, we also analyze average net earnings for the treatments with and without reward opportunities. Averaging across all 25 periods, net profits are 24.55 tokens in the VCMs with reward opportunities and 23.27 tokens in the VCMs without reward opportunities. This difference is not statistically significant (ranksum Wilcoxon test, $p = 0.452$). Nor is any significant difference observed when we look at average net earnings in either the first 5 periods or the last 10 periods.

These results are supported when we examine how the effectiveness of the punishment and reward institutions compares with the standard VCM institution. In particular, we calculate the average earnings in each period between each institution and the VCM institution. We normalize this difference by the average earnings in the VCM institution in the same period. The evolution of the relative earnings is shown in Figure 2. We observe that, in the short run, institutions with punishment yield lower average earnings relative to the standard VCM institution. However, the benefits from higher contributions

tend to outweigh the costs from assigning punishment, resulting in higher average earnings over time. Regarding the VCM with reward, we observe that average earnings are similar to those in the VCM institution as they fluctuate around the zero level (which indicates no difference in average earnings with respect to the VCM institution).[16]

Figure 2: Average relative earnings in each period



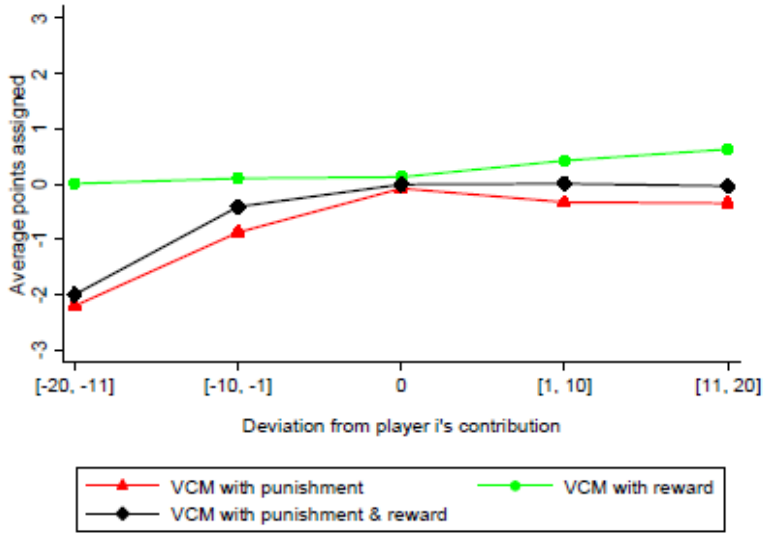
3.2 Assigning points for sanctions and rewards

This section analyzes behavior in the second stage of the VCM treatments by discussing how sanctions and rewards are actually used. Figure 3 depicts how subjects mete out punishments and rewards based on how much the peer's contribution deviates from the punisher's/donor's contribution. The vertical axis indicates the average points assigned to a group member by player *i*. The horizontal axis indicates the deviation in discrete intervals of the recipient's contribution from the contribution of the punisher/donor (player i). For example, a subject in the VCM treatment with punishment assigned, on average, –2.21 points to those who contributed between 11 and 20 tokens less than him/her.[17]

[16] Interestingly, we also record end game effects, especially in the VCM institutions which allow for punishment opportunities.
[17] The actual points assigned by the punisher/donor in each deviation interval are shown in Table D.4 in Appendix D.

Figure 3 provides evidence that in both treatments where punishment opportunities exist, negative deviations from the punisher are strongly sanctioned. In particular, the greater the negative deviation is from the punisher's contribution, the harsher the punishment. Not surprisingly, in the VCM treatment with reward, we observe that negative deviations are not rewarded. Positive deviations (using the donor's contribution as a base) are rewarded, with the reward being increasing in the size of the deviation. However, for the VCM treatment with punishment and reward, the average points assigned for positive deviations are half as much as the average points assigned in the reward only treatment.

Figure 3: Average Points Assigned as a Function of Deviation
from the Sanctioning/Rewarding Player's Contribution



To analyze the characteristics of those subjects who assign sanctions and/or rewards, we employ a random effect Tobit model. The dependent variable is the costs an individual incurs by assigning sanctions and/or rewards in a given period, while the explanatory variables include the standardized preference measures and the absolute negative and positive deviations from player *i*'s contribution as negative/positive deviations elicit different punishment/reward responses. Our regression results are presented in Table 2.

Table 2: Preference Measures and Assignment of Points: Regression Results

	Dependent Variable: Points assigned by player i		
	VCM with punishment	**VCM with reward**	**VCM with punishment and reward**
Standardized risk aversion	0.16***	–0.07**	–0.03
	[0.05]	[0.03]	[0.03]
Standardized loss aversion	0.09*	0.13***	0.13***
	[0.05]	[0.04]	[0.03]
Standardized ambiguity (without loss) aversion	0.05	0.06	–0.01
	[0.05]	[0.04]	[0.03]
Standardized ambiguity (with loss) aversion	–0.07	–0.02	–0.01
	[0.06]	[0.04]	[0.03]
Switcher?	0.11	0.26***	0.17***
	[0.13]	[0.08]	[0.05]
Absolute negative deviation	–0.13***	–0.01***	–0.13***
	[0.003]	[0.002]	[0.004]
Positive deviation	–0.01**	0.05***	–0.002
	[0.003]	[0.002]	[0.004]
Constant	–0.17***	0.09***	–0.05
	[0.04]	[0.03]	[0.03]
Observations	3,600	3,600	3,600

Note: Random effect Tobit regressions with standard errors in square brackets. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level, and *** at the 1-percent level.

In line with previous results in the literature, we find that absolute negative deviations are significantly and negatively correlated with the points assigned for punishments and/or rewards in all three treatments. This result implies that the further a subject falls below the punisher's/donor's contribution, the more negative points are assigned to him/her. In addition, positive deviations are significantly positively correlated with points assigned in the VCM treatments with reward and with punishment, implying that the more a subject positively deviates from the donor's contribution the more rewards the donor assigns to him/her, and analogously (with a much smaller magnitude) for punishment.

Furthermore, we document statistically significant relationships between preference measures and punishment/reward behavior.[18] Specifically, standardized risk aversion is positively (negatively) correlated with the assignment of points in the VCM treatment with punishment (reward), while loss aversion is significantly and positively correlated with assignment of points in all three treatments. The more risk averse a subject is, the fewer punishment points he/she assigns and the more rewards he/she assigns; whereas, the more loss averse a subjects is, the more expenditure he/she makes (by assigning more points). As subjects may be unaware of the overall efficiency of the VCM institution with punishment, a risk averse subject may be willing to hedge and thus assign fewer costly points, especially since the impact from receiving points is higher than the cost. We also find that risk aversion is negatively correlated with the assignment of reward points, suggesting that subjects may not be certain about how rewards will work and, as a result, they are more reluctant to use them. When punishments and rewards are combined, we find that loss aversion is a significant determinant for assigning points, which is in line with the previous observation that loss aversion explains the assignment of reward points On the other hand, we do not observe significant links between ambiguity preferences and reward/punishment behavior, at least when all preferences appear in the same regression.[19]

Separately we find that being a switcher is positively and significantly correlated with the assignment of points in the VCM treatment with reward and in the VCM treatment with punishment and reward. This may be an indication that these subjects are more active or impulsive participants when they play the public goods games, which in turn may explain why they were switchers in the preference elicitation tasks.

RESULT 1: *Preference measures significantly affect the points assigned in all three treatments.*

As a final step, we check to see whether individuals' underlying preferences over institutions involving reward or punishment are linked to their ultimate use of reward and/or punishment. The results of this analysis are reported in Table 3. In addition to the effect of deviations from one's own contribution level, as reported above, the most striking finding is that those subjects who said they preferred a reward environment actually give out

[18] Our results remain robust when we control for demographics (available upon request).
[19] Ambiguity with losses is significant in all treatments when it is the only preference measure used, but we believe the joint regression is the stronger and more theoretically appropriate specification.

significantly more reward points. Recall that this result is not a selection effect, since everyone was equally likely to have ended up in the reward setting, and indeed the subjects are aware that they are matched with people with all different possible preferences. Nevertheless, these subjects seem to have an underlying belief in the efficacy of rewards to induce positive behavior (or simply for its own sake). No similar link is found in the case of punishment.

Table 3: Institutional Preferences and the Assignment of Points for Punishment and Reward

	Dependent Variable: Points assigned by player i		
	VCM with punishment	**VCM with reward**	**VCM with punishment and reward**
Absolute negative deviation	–0.13***	–0.01***	–0.13***
	[0.004]	[0.002]	[0.004]
Positive deviation	–0.01**	0.05***	0.002
	[0.004]	[0.002]	[0.004]
Switcher	0.15	0.18**	0.14**
	[0.27]	[0.08]	[0.06]
Ranking of VCM with punishment	0.02		
	[0.04]		
Ranking of VCM with reward		–0.05*	
		[0.03]	
Ranking of VCM with punishment and reward			0.01
			[0.02]
Constant	–0.25**	0.23***	–0.05
	[0.11]	[0.07]	[0.07]
Observations	3,600	3,600	3,600

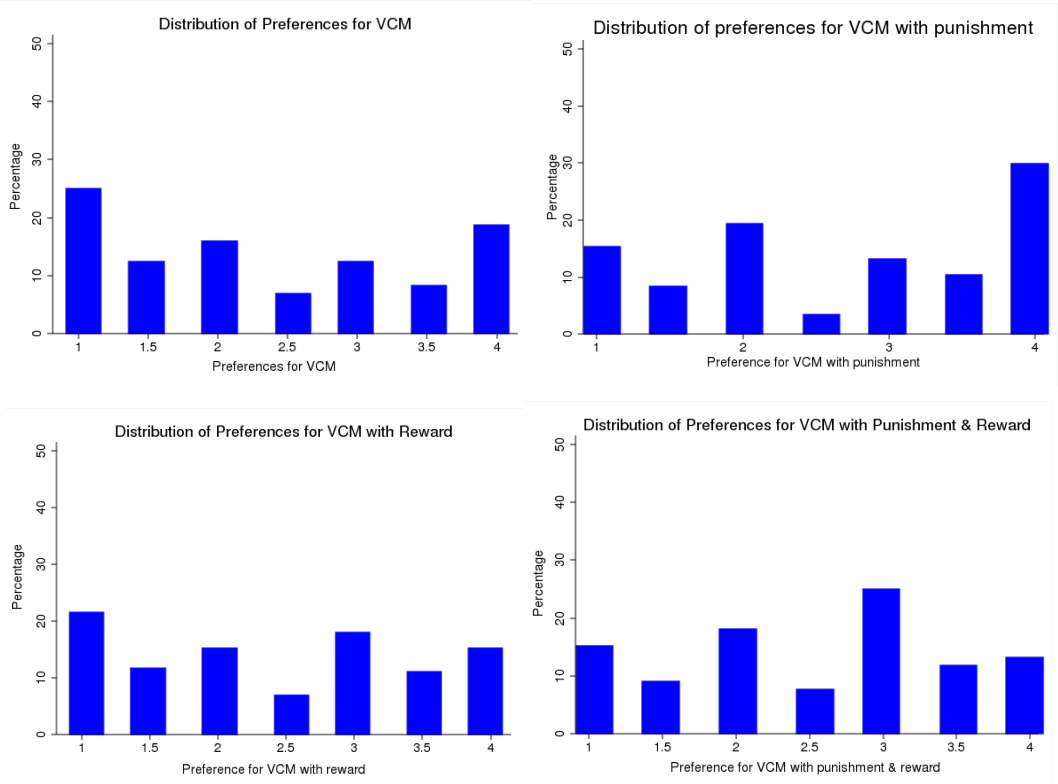
Note: Random effect Tobit regressions with robust standard errors in square brackets. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level, and *** at the 1-percent level.

2. Choice of institutions

Figure 4 displays the subjects' distribution of preferences for each of the four institutions. The horizontal axis in each panel indicates the ranked preference for each institution using a four-point scale, where "1" denotes a subject's most favored institution and "4" denotes the least favored institution. In case of a tie between two or more institutions, we assigned the average value of the ranked positions that the two institutions occupied. This

implies that the sum of the rankings of each institution for a given subject is always equal to 10.

Figure 4: Distribution of Preferences Over Institutions



Note: 1= "Most preferred," 4= "Least preferred"

Most subjects exhibit a preference to participate in institutions which do not include punishment opportunities. As Figure 4 suggests, the VCM treatment is assigned a ranking of 1, 1.5, or 2 by 53.65 percent of subjects, while the corresponding percentage of subjects who rank the VCM with reward treatment as 1, 1.5, or 2 is 51.04 percent. On the other hand, the VCM with punishment and reward, and the VCM with punishment treatments are ranked 1, 1.5, or 2 by 43.75 percent and 40.10 percent of subjects, respectively.[20]

A relevant question to address is what influences the amount an individual is willing to pay (or would need to be paid) for each institution. Table 4 contains regression models, with two models corresponding to each institution. In models 1, 3, 5, and 7, we examine the extent to which a subject's risk, loss, and ambiguity preferences affect preferences over each

[20] As a complementary measure of preferences, we explored how much each subject was willing to pay/get paid as a percentage of the maximum amount allowed (that is, £5). This analysis conveys a similar message to our earlier discussion. Specifically, 60.42 percent and 53.13 percent of subjects report that they want to pay a non-negative amount to participate in the VCM treatment and the VCM with reward treatment; while the corresponding percentages for the VCM with punishment and reward and the VCM with punishment treatments are 52.6 percent and 43.75 percent, respectively.

institution separately; while in models 2, 4, 6, and 8, we check the robustness of these results to the inclusion of more explanatory variables including overconfidence levels, the number of correct CRT questions, gender, age, and nationality. Our preference measures have been standardized to have a mean of zero and a standard deviation of one.

Table 4: Preference Measures and Choice of Institutions

	Dependent variable: Intensity of preferences (measured as the amount stated for each institution)							
	VCM		**VCM with punishment**		**VCM with rewards**		**VCM with punishment & rewards**	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Risk Aversion	–8.84	–11.32	11.11	11.92	–12.15	–13.41	9.88	12.81
	(19.58)	(19.68)	(17.92)	(17.64)	(16.57)	(16.48)	(16.71)	(16.93)
Loss Aversion	4.89	–0.49	12.85	9.87	–8.89	–6.43	–8.85	–2.95
	(20.41)	(20.80)	(23.05)	(23.94)	(20.73)	(21.08)	(19.40)	(19.52)
Ambiguity Aversion	17.05	19.19	–11.18	–8.57	–4.25	–4.45	–1.62	–6.17
	(18.48)	(19.93)	(19.70)	(20.74)	(19.95)	(21.11)	(17.68)	(18.45)
Ambiguity (with loss) Aversion	–34.87	–30.20	0.68	3.33	8.74	7.90	25.45	18.98
	(22.46)	(22.35)	(23.01)	(23.61)	(21.21)	(21.44)	(21.96)	(22.37)
Switcher	–25.00	–20.70	–22.40	–38.59	9.77	16.27	37.63	43.02
	(37.17)	(38.22)	(33.48)	(34.20)	(32.85)	(33.98)	(34.72)	(34.07)
Overconfidence		–38.98		75.66		9.49		–46.18
		(94.32)		(101.68)		(93.79)		(82.27)
Correct CRT questions		0.03		–35.07**		21.29		13.74
		(15.78)		(15.53)		(14.33)		(13.41)
Male		–39.14		25.83		–37.70		51.01
		(36.99)		(40.00)		(32.90)		(33.06)
Age		–4.22		1.99		–0.74		2.98
		(4.50)		(4.97)		(4.41)		(4.33)
UK nationality		72.68*		–13.86		22.49		–81.32**
		(37.02)		(39.80)		(37.36)		(35.44)
Constant	21.63	124.00	–26.92	–61.43	–10.62	–30.36	10.71	–32.21
	(16.10)	(121.45)	(19.50)	(132.19)	(17.86)	(126.87)	(17.16)	(118.41)
Observations	192	192	192	192	192	192	192	192

Note: OLS regressions with robust standard errors in square brackets. The preference measures reported in the table are standardized. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level, and *** denotes significance at the 1-percent level.

Our regression analysis shows that individual preference measures are not good predictors of institutional choice, either in terms of magnitude or statistical significance. This result is robust to several alternate specifications: excluding the 40 subjects who switched back and forth in at least one preference elicitation task (see Table D.3 in the Appendix); using the ordinal rather than the cardinal strength of intensity over institutional environments; and including only one individual preference at a time in the regression (potentially preferable because of positive correlations between some of those measures). We do find that some institutional preferences are partially predicted by cognitive self-control (as assessed by the CRT task), which warrants further study. So although we and others (based on previous work as described in the introduction) predicted that risk, loss, and ambiguity attitudes would be important factors at this stage, that does not appear to be the case in our data.

To further explore this possible relationship, one approach is to look for heterogeneity within the sample. In particular, note that stating a preference intensity of 500 pence for a particular institution implies an unrealistically large per-period gain in profits and may be an indicator that a given subject did not fully understand the situation.[21] If we restrict attention only to those subjects who expressed preference intensities in either direction of 200 pence at most (there are 135 such subjects from our total of 192), we find instead that being more risk averse and being more ambiguity averse both significantly predict a preference for the default VCM institution, meaning the simplest environment with no punishment or reward mechanisms, as would be expected. This result is robust to the specific cutoffs on intensity (within a reasonable range), but since this expectation was not our specific hypothesis in advance we hesitate to give it undue weight. Nevertheless, this result suggests that at least for some individuals, their attitude toward uncertainty is indeed an important determinant for their preference toward social structure.

RESULT 2: *Overall, preference measures do not seem to matter for institutional choice. However, further work is warranted to explore the boundaries and extent of this relationship.*

3. Preference measures, cognitive reflection test (CRT) questions, and demographics

We next turn our attention on whether and if so, how preferences are interrelated to each other and correlated with cognitive self-control (as measured by subjects' responses in a cognitive reflection test). In the first part of our experiment, we elicited subjects' risk, loss,

[21] We thank Urs Fischbacher for suggesting this line of reasoning.

and ambiguity preferences. Recall that for a given preference measure, subjects had to make seven separate choices between a fixed amount safe option and a lottery, in which the payment amount increased from row to row moving down the table. We use the number of times a subject chose the safe option as a measure of his/her attitudes corresponding to the specific preference measure. For example, when eliciting risk preferences, never choosing the risky option indicates extreme risk aversion, whereas choosing seven risky options indicates extreme risk-seeking behavior. Table 5 presents summary statistics on how often the subjects chose the sure payoff for each preference measure. We observe that, on average, our subjects are ambiguity averse with respect to both gains and losses. Performing a Wilcoxon matched-pairs signed rank test, we find that the difference between the mean values of risk aversion and ambiguity (without loss) aversion is highly statistically significant ($p = 0.000$). This is also the case when we compare the mean value of loss aversion and that of ambiguity (with losses) aversion ($p = 0.000$). Comparing the mean values of risk aversion and loss aversion, we find significant differences at the 5-percent level ($p = 0.038$), with more risk aversion than loss aversion – an unusual but not unprecedented result.

Table 5: Means and Standard Deviations of Preference Measures

Preference measure	**Mean**	**Standard deviation**
Risk aversion	3.51	1.65
Loss aversion	2.99	1.97
Ambiguity aversion (without losses)	4.28	1.71
Ambiguity aversion (with losses)	3.63	2.22

As is always observed with this elicitation style for preference measures, some of our subjects switch more than once between the safe option and the lottery option, a choice which is considered to be inconsistent behavior. As mentioned earlier, we refer to these subjects as "switchers." In our sample, there are 40 subjects who switched more than once in at least one preference measure. After excluding these 40 subjects, we find similar mean numbers of safe choices for each preference measure (see Table D.1 in the Appendix D for summary statistics). The differences in mean values, documented earlier, are robust to this exclusion,

with the exception that the mean value of risk aversion and that of loss aversion is now significantly different at the 10-percent level (p = 0.052).

We next examine whether preference measures are correlated with each other. As Table 6 suggests (p-values are reported in square brackets), we find significant correlations between all pairs, except for the dyad of loss aversion and ambiguity aversion (without losses). In particular, the positive signs of the reported coefficients indicate that the more risk averse a subject, the more averse they are to loss and ambiguity.[22] In addition, loss averse subjects are more ambiguity averse only when losses are involved, whereas the more ambiguity averse a subject is toward gains, the more ambiguity averse he/she is to losses. These conclusions are robust when the 40 switchers are excluded (see Table D.2 in the Appendix D). Interestingly, previous studies that have used different instruments from ours do not find significant correlations between preferences, suggesting that their findings are "inconsistent with the notion that individuals have a fixed, domain-general utility function that is applicable to all risky situations" (Eckel and Wilson 2004, p. 457). We view our evidence on correlations as a means of validating our preference measures, which are fairly standard but have not yet been replicated across a fully diverse set of environments. Of course, it is difficult to know for certain which, if any, of the preference measures are most closely capturing the underlying constructs.

Table 6: Pair-wise Pearson Correlation Coefficients of Preference Measures

	Risk aversion	**Loss aversion**	**Ambiguity aversion (without losses)**	**Ambiguity aversion (with losses)**
Risk aversion	1			
Loss aversion	0.25*** [.00]	1		
Ambiguity aversion (without losses)	0.40*** [.00]	0.09 [.21]	1	
Ambiguity aversion (with losses)	0.16** [.02]	0.61*** [.00]	0.25*** [.00]	1

[22] The latter result matches survey data reported in Butler, Guiso, and Japelli (2011).

We next explore how the level of individual cognition and executive function (as measured by the number of correct CRT questions) is related to subjects' preferences over risk, loss, and ambiguity. To address this question, we employ an ordered probit analysis where the dependent variable indicates the number of correct CRT questions that a subject provided. Table 7 presents our three regression models. In Model 1, we include whether a subject was a switcher and the four standardized preference measures as explanatory variables. Model 2 checks for the robustness of these results by adding age, gender, nationality, and overconfidence. Model 3 includes a number for other controls such as father's education, economics or business major, perceptions of fairness, whether a subject has ever participated in an economics experiment, number of other participants a subjects knows, political party, and religion affiliation.

Table 7: CRT Questions: Regression Results

	Dependent variable: # correct CRT questions		
	1	**2**	**3**
Switcher	–0.56***	–0.51**	–0.48**
	[0.21]	[0.23]	[0.23]
Standardized Risk aversion	–0.02	0.00	–0.02
	[0.09]	[0.09]	[0.09]
Standardized Loss aversion	–0.21**	–0.16	–0.17
	[0.11]	[0.11]	[0.11]
Standardized Ambiguity (w/o loss) aversion	0.10	0.08	0.08
	[0.09]	[0.09]	[0.09]
Standardized Ambiguity (w/ loss) aversion	0.14	0.11	0.14
	[0.11]	[0.10]	[0.10]
Male		0.31*	0.32*
		[0.17]	[0.17]
Age		0.00	0.00
		[0.02]	[0.02]
UK nationality		–0.136	–0.08
		[0.17]	[0.19]
Overconfidence		1.100**	1.03**
		[0.50]	[0.52]

Controls for other demographics?	No	No	Yes
Observations	192	192	192

Note: Ordered probit with robust standard errors reported in square brackets. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level, and *** at the 1-percent level.

Our regression analysis suggests that loss averse subjects appear to correctly answer fewer CRT questions. However, this effect vanishes when we control for the other demographic characteristics we collected. Interestingly, these characteristics have a significant impact on our specific measure of subjects' cognitive abilities. In particular, we find that males tend to answer more CRT questions correctly and subjects who report high overconfidence levels also tend to answer more CRT questions correctly.[23] A noteworthy aspect of our regression results has to do with the subjects who switched back in at least one of our preference elicitation tasks. In all three models, it turns out that "switchers" tend to answer fewer CRT questions correctly. The main findings from this section are summarized in Result 3.

RESULT 3: *Preference measures are positively and significantly correlated with each other. An individual's cognitive executive function, as measured by the CRT, are related to his or her degree of loss aversion, gender, overconfidence, and whether the subject switched more than once in at least one preference elicitation task.*

V. Concluding remarks

This paper provides experimental evidence that a subject's individual preferences are significantly related to his/her revealed economic preferences, as measured by observing behavior in public goods institutions. Somewhat surprisingly, this link is observed only for actions taken in the games themselves and not also for the initial revealed preferences over

[23] The gender differences we observe in relation to correct CRT questions replicate the findings of the seminal paper by Frederick (2005), suggesting that a possible explanation for this could be that men are more likely to reflect on their answers and less likely to choose intuitive responses compared to women. A similar gender effect on CRT responses has also been found by Oechssler, Roider and Schmitz (2008). This latter paper also explores the relationship between cognitive ability (as measured by correct CRT questions) and overconfidence (as measured by subjects' estimates about the number of CRT questions they had solved correctly). They document a positive and significant relationship between these two variables: those who answer more CRT questions correctly judge their performance more accurately. As the authors state in their paper (page 6): "this has to do with the property of the CRT questions that induce impulsive decision makers to judge their intuitive but wrong answers as correct."

the incentive structure of institutions. The set of four institutions we are concerned with in this study are: the standard VCM, the VCM with punishment, the VCM with reward, and the VCM with punishment and reward. Our novel experimental design also provides us with a rich dataset which allows us to analyze whether and how preference measures are correlated with each other, as well as whether these measures gauge the impact of social preferences such as negative or positive reciprocity (as reflected by punishing and rewarding behavior). These issues have been relatively less explored in the literature and our experiment provides a complete set of answers to these questions in the context of a public goods game.

Our paper also demonstrates the significance of individual preference measures to economic behavior. Most studies (e.g., Borghans et al. 2009; Dohmen et al. 2011; Fréchette, Schotter, and Trevino 2011; Butler, Guiso, and Jappelli 2011) limit their attention to particular aspects of individuals' preferences. We extend this literature by examining four preference measures: risk aversion, loss aversion, ambiguity aversion, and ambiguity (with loss) aversion. We show that our preference measures are strongly correlated with each other, implying that individuals have a general utility function that is applicable to situations involving risk and ambiguity. In addition, we find that executive function and cognitive fluidity, as measured by the number of correct answers to CRT questions, is also associated with preference measures.

The origin of preferences over institutions is important as they can have serious implications for a society's evolution, as well as on its economic performance and welfare over time. It is therefore important to analyze from an empirical perspective how institutions are determined. Our findings suggest that most people at least initially prefer to participate in institutions where no sanctioning is present. We also show that the demand for these environments is affected by cognitive sophistication but importantly does not seem to be affected by attitudes toward risk, loss, or ambiguity.[24] Given that there is a natural link between e.g. loss aversion and the existence of negative sanctions, and given that experimentally assessed risk and ambiguity preferences have correlated with behaviors in other contexts (although not always), this is an unexpected and interesting result.[25] One possible explanation is that strategic concerns overrode individual preferences in this finitely-repeated and highly structured setting with monetary payoffs. Other explanations could

[24] As reported above, we do see some interesting links when the sample is restricted to subjects who are not at the tails of the distribution. Since this was an *ex post* analysis, we cannot be as statistically confident in it. However, it is reassuringly suggestive for future study.

[25] Although note that some previous literature (see Allen and Lueck 1995 for a review) similarly concludes that there is a weaker relationship between risk attitudes and contract choice than has been hypothesized by many researchers.

include the fact that participants had not directly experienced any of the environments beforehand (which was a deliberate design choice) or that we did not focus on social preferences. Both of these ideas are natural avenues for further study.

We document significant relationships between reciprocity (whether driven by strategic concerns or moral disgust) and preference measures. For instance, risk aversion is positively (negatively) correlated with points assigned in the VCM with punishment (VCM with reward), whereas loss aversion is positively related with point assignment in the VCM with punishment and reward. Our paper confirms and validates previous findings that although in the short run institutions with sanctions yield lower welfare, in the long run institutions are more efficient with a punishment mechanism: subjects' earnings, on average, are higher relative to the average earnings of those who participate in institutions without sanctioning options.

Our evidence also suggests some new avenues to enrich the economic theory on social interactions and the emergence of institutions. For example, recent social preference models can explain contribution decisions and punishment patterns in public good environments but lack the ability to incorporate factors such as preference measures, which have been shown to have significant predictive behavior. Our research provides further evidence that risk and social preferences should be incorporated into economic analysis, as these are related notions that help us further understand certain aspects of economic behavior.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91(5): 1369–1401.

Acemoglu, Daron, Simon Johnson, James A. Robinson. 2002. "Reversal of Fortunes: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics* 117(4): 1231–1294.

Ackerberg, Daniel A., and Maristella Botticini. 2002. "Endogenous Matching and the Empirical Determinants of Contract Form." *Journal of Political Economy* 110(3): 564–591.

Allen, Douglas W., and Dean Lueck. 1995. "Risk Preferences and the Economics of Contracts." *American Economic Review, Papers and Proceedings* 85(2): 447–451.

Bartling, Björn, Ernst Fehr, Michel André Maréchal, and Daniel Schunk. 2009. "Egalitarianism and Competitiveness." *American Economic Review, Papers and Proceedings* 99(2): 93–98.

Binswanger, Hans P. 1980. "Attitudes toward Risk: Experimental Measurement in Rural India." *American Journal of Agricultural Economics* 62(Aug): 395-407.

Borghans, Lex, Bart. H. H. Golsteyn, James J. Heckman, and Huub Meijers. 2009. "Gender Differences in Risk Aversion and Ambiguity Aversion." *Journal of the European Economic Association* 7(2–3): 649–658.

Botelho, Anabela, Glenn W. Harrison, Lígia M. Costa Pinto, and Elizabet E. Rutström. 2005. "Social Norms and Social Choice." Núcleo de Investigação em Microeconomia Aplicada (NIMA) Working Paper 30. Braga: University of Minho. Available at http://www3.eeg.uminho.pt/publications/NIMAwp30.pdf.

Butler, Jeffrey V., Luigi Guiso, and Tullio Jappelli. 2011. "The Role of Intuition and Reasoning in Driving Aversion to Risk and Ambiguity." CEPR Discussion Paper 8334. London: Centre for Economic Policy Research. Available at http://www.cepr.org/pubs/new-dps/dplist.asp?dpno=8334.asp.

Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton, NJ: Princeton University Press.

Chaudhuri, Ananish. 2011. “Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature.” *Experimental Economics* 14(1): 47–83.

Cheung, Steven N.S. 1969. “Transaction Costs, Risk Aversion, and the Choice of Contractual Arrangements.” *Journal of Law and Economics* 12(1): 23–42.

Dal Bó, Pedro. 2010. “Experimental evidence on the workings of democratic institutions”, Working Paper.

Dal Bó, Pedro, Andrew Foster, and Louis Putterman. 2010. “Institutions and Behavior: Experimental Evidence on the Effects of Democracy.” *American Economic Review* 100(5): 2205–2229.

de Quervain, Dominique J.-F., Urs Fischbacher, Valerie Treyer, Melanie Schellhammer, Ulrich Schnyder, Alfred Buck, and Ernst Fehr. 2004. “The Neural Basis of Altruistic Punishment.” *Science* 305(5688): 1254–1258.

Dohmen, Thomas, and Armin Falk. 2011. “Performance Pay and Multidimensional Sorting: Productivity, Preferences and Gender.” *American Economic Review* 101(2): 556–90.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp and Gert G. Wagner. 2011. “Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences.” *Journal of the European Economic Association* 9(3): 522–550.

Easterly, William, and Ross Levine. 2003. “Tropics, Germs, and Crops: How Endowments Influence Economic Development.” *Journal of Monetary Economics*. 50(1): 3–39.

Eckel, Catherine C., and Rick K. Wilson. 2004. “Is Trust a Risky Decision?” *Journal of Economic Behavior and Organization* 55(4): 447–465.

Ertan, Arhan, Talbot Page, and Louis Putterman. 2009. “Who to Punish? Individual Decisions and Majority Rule in Mitigating the Free Rider Problem.” *European Economic Review* 53(5): 495–511.

Fehr, Ernst and Klaus M. Schmidt. 2006. “The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories”, Handbook on the Economics of Giving, Reciprocity and Altruism, Vol. 1, 615-691 edited by Serge-Christophe Kolm and Jean Mercier Ythier, Amsterdam: Elsevier.

Fréchette, Guillaume R., Andrew Schotter, and Isabel Trevino. 2011. "Personality and Choice in Risky and Ambiguous Environments: An Experimental Study." Department of Economics Working Paper. New York: New York University. Available at https://files.nyu.edu/gf35/public/print/Frechette_2011c.pdf.

Frederick, Shane. 2005. "Cognitive Reflection and Decision Making" *Journal of Economic Perspectives* 19(4): 25–42.

Gächter, Simon and Benedikt Herrmann. 2005. "Human Cooperation from an Economic Perspective." In *Cooperation in Primates and Humans Mechanisms and Evolution*, ed. By Peter M. Kappeler and Carel P. van Schaik, pp. 275–302. Heidelberg: Springer-Verlag.

Gächter, Simon and Benedikt Herrmann. 2009. "Reciprocity, Culture and Human Cooperation: Previous Insights and a New Cultural Experiment." *Philosophical Transactions of the Royal Society B – Biological Sciences* 364(1518): 761–806.

Gächter, Simon, Elke Renner, and Martin Sefton. 2008. "The Long-Run Benefits of Punishment." *Science* 322(5907): 1510.

Greiner, Ben. 2004. "An Online Recruitment System for Economic Experiments." In *Forschung und wissenschaftliches Rechnen*. Gesellschaft für wissenschaftliche Datenverarbeitung Bericht, Vol. 63, ed. By Kurt Kremer and Volker Macho, pp. 79–93. Göttingen, Germany: Gesellschaft für Wissenschaftliche Datenverarbeitung.

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach. 2006. "The Competitive Advantage of Sanctioning Institutions." *Science* 312(5770): 108–111.

Gürerk, Özgür, Bernd Irlenbusch and Bettina Rockenbach. 2009. "Voting with Feet: Community Choice in Social Dilemmas." IZA Discussion Paper 4643. Bonn, Germany: Institute for the Study of Labor. Available at http://ftp.iza.org/dp4643.pdf.

Holt, Charles A., and Susan. K Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92(5): 1644–1655.

Humphrey, Steven J., and Elke Renner. 2011. "The Social Costs of Responsibility." Centre for Decision Research & Experimental Economics (CeDEx) Discussion Paper 2011–02. Nottingham, UK: University of Nottingham. Available at http://www.criticaltheology.ac.uk/cedex/documents/papers/2011-02.pdf.

Kamei, Kenju. 2011." Democracy and Resilient Pro-Social Behavioral Change: An Experimental Study." Department of Economics Working Paper. Providence, RI: Brown University. Available at http://www.econ.brown.edu/students/kenju_kamei/jmp.pdf.

Kocher, Martin G., Peter Martinsson, Dominik Matzat, and Conny Wollbrant. 2011. "The Role of Beliefs, Trust, and Risk in Contributions to a Public Good." Working Papers in Economics 482. Gothenberg, Sweden: University of Gothenburg. Available at http://hdl.handle.net/2077/24123.

La Porta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert W. Vishny. 1998. "Law and Finance." *Journal of Political Economy* 106(6): 1113–1155.

Ledyard, John. 1995. "Public Goods: A Survey of Experimental Research." In *Handbook of Experimental Economics*, ed. By John H. Kagel and Alvin E. Roth, pp. 111–194. Princeton, NJ: Princeton University Press.

Magreiter, Magdalena, Matthias Sutter and Dennis Dittrich, 2005. "Individual and collective choice and voting in common pool resource problem with heterogeneous actors." *Environmental and Resource Economics* 32(2): 241–271.

North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge, UK: Cambridge University Press.

Noussair, Charles, and Fangfang Tan. 2011. "Voting on punishment systems within a heterogeneous group." *Journal of Public Economic Theory*, 13(5): 661–693.

Oechssler, Jörg, Andreas Roider and Patrick W. Scmhitz. (2008). "Cognitive abilities and behavioral biases" *IZA Discussion Paper No. 3481*.

Ostrom, Elinor, James Walker, and Roy Gardner. 1992. "Covenants With and Without a Sword: Self-Governance is Possible." *American Political Science Review* 86(2): 404–417.

Putterman, Louis, Jean-Robert Tyran, and Kenju Kamei. 2011. "Public Goods and Voting on Formal Sanction Schemes." *Journal of Public Economics* 95(9–10): 1213–1222.

Schechter, Laura. 2007. "Traditional Trust Measurement and the Risk Confound: An Experiment in Rural Paraguay." *Journal of Economic Behavior and Organization* 62(2): 272–292.

Sutter, Matthias, Stefan Haigner, and Martin G. Kocher. 2010. “Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations.” *Review of Economic Studies* 77(4): 1540–1566.

Tabellini, Guido. 2008. “Presidential Address: Institutions and Culture.” *Journal of the European Economic Association* 6(2–3): 255–294.

Tyran, Jean-Robert, and Lars P. Feld, 2006. “Achieving compliance when legal sanctions are non-deterrent”, *Scandinavian Journal of Economics* 108(1): 135–156.

Weinhold, Diana, and Paul J. Zak. 2005. “The Choice of Institutions: The Role of Risk and Risk-Aversion.” Working Paper. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=784765.

Yamagishi, Toshio. 1986. “The Provision of a Sanctioning System as a Public Good.” *Journal of Personality and Social Psychology* 51(1): 110–116.